Understanding Choice Behavior Beyond Option Scaling Using Structural Equation Models[*]

Rung-Ching Tsai[1] and Ulf Böckenholt[2]
[1]*National Taiwan Normal University and* [2]*Northwestern University*

Abstract: Random Utility models have been shown useful in scaling choice options, as well as in providing a rich source of information about individual differences and perceived similarity relationships among choice alternatives. Modeling of preference data such as rankings was made easier by representing utilities as latent factors in a structural equation modeling framework. In this paper, we extend such an SEM approach to analyze ranking data and other types of ordinal data simultaneously. This combination of both absolute and relative judgment data can enrich our understanding of individual differences in multiple domains including preference and attitude.

Key words: Ranking data, random utility models, structural equation models, scaling.

1. Introduction

Starting from Thurstone's (1927) seminal work on discriminal process in psychophysical experiments, the task of paired comparisons and ranking have often been adopted to find scaling values of the choice stimuli. In a ranking task, all items are presented altogether and a judge is asked to provide orders or ranks of all the choice items.

It has been shown that by postulating the underlying mechanism as a utility maximization for ranking and paired comparisons, utility models not only arrive at values which characterize the mean preferences of the choice items, but also allow for the possibility of understanding the individual difference in their judgmental process. The random utility paradigm has been highly influential in the development of many preference models, especially in the econometric and transportation literature (for reviews, see Train, 2003; McFadden, 2001).

By focusing on the mean structure of choice items, one might overlook the equally important information about individual difference embedded in the variability of comparative judgments (Bock, 1958; Takane, 1987; Böckenholt & Tsai, 2001). It can be seen that, in most applications involving paired comparison and ranking tasks nowadays, scaling the stimuli is still their main purpose and the simplest Case V Thurstonian model is commonly assumed perhaps for computational reason (Furuya, et al., 2005; Kojima & Kusumi, 2006; Prietoa & Alonsoa, 2000). However, it has been shown that disregard for the dependencies between the latent utilities and systematic individual differences is a serious model misspecification that leads to both incorrect statistical and substantive conclusions (Bock, 1958; Takane, 1987). Estimation of the Thurstonian choice models other than the simple Case V have been made possible from a number of past work on estimation issues (Takane, 1987; Böckenholt & Tsai, 2001; Maydeu-Olivares, 1999 2002; Yao & Böckenholt, 1999; Yu, 2000). However, most of earlier work on estimation required statistical expertise and failure to implement those approaches into standard software hindered the applications of such models.

Maydeu-Olivares and Böckenholt (2005) gave a detailed account on the formulation, identification and estimation of both paired comparison and ranking models under a structural equation (SEM) modeling framework and suggested a flow chart which could be used as a strategy for model selection. By embedding Thurstonian choice models under the SEM framework, there is no doubt that such an approach will facilitate not only the number of future applications, but also more further extensions with the advances of SEM techniques. In particular, it is shown that ranking data could be converted into multiple binary responses and consequently could be analyzed using SEM models suitable for dichotomous items to ease estimation difficulty that the researchers often encounter while analyzing ranking data.

The purpose of this article is to extend Maydeu-Olivares and Böckenholt's (2005) approach in analyzing ranking data alone to include covariates and other measures of latent traits of the judges in the analysis as well. By modeling ranking and all the other variables in the data within an SEM framework, we are able to gain our understanding, from the estimates of the model parameters, in individual differences in the utilities the judges assigned to the options in the ranking task, as well as in how these utilities relate to the covariates or some personality traits of the judges.

2. Model Description

2.1 Thurstonian ranking models

In a complete ranking task, choice items are presented at once and the judges

are asked to rank or order all the items. For example, when a judge is asked to give a complete ranking of all four items in the choice set (j, k, l, m), the ranking pattern of $(m \succ k \succ j \succ l)$ indicates that item m is the most, item k the second, item j the third, and item l the least preferred, where $A \succ B$ indicates that option A is preferred to option B. For each ranking pattern, there exists a response pattern of multiple paired comparisons which corresponds to the same preference ordering of the choice items. For instance, the ranking pattern of $(m \succ k \succ j \succ l)$ corresponds to the pairwise outcomes of $((j, k), (j, l), (j, m), (k, l), (k, m), (l, m)) = (0, 1, 0, 1, 0, 0)$. According to Thurstonian ranking models, the choice between any two items j and k for judge i is determined by the difference in his or her perceived utilities y_{ijk} formulated as follows that

$$y_{ijk} = \eta_{ij} - \eta_{ik}, \tag{1}$$

where η_{ij} and η_{ik} represent the perceived utilities of items j and k for judge i, respectively. In other words, for ranking of the options j, k, l, and m, we would have

$$\mathbf{y}_i = \begin{pmatrix} y_{ijk} \\ y_{ijl} \\ y_{ijm} \\ y_{ikl} \\ y_{ikm} \\ y_{ilm} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \eta_{ij} \\ \eta_{ik} \\ \eta_{il} \\ \eta_{im} \end{pmatrix} = \mathbf{A}\,\boldsymbol{\eta}_i \tag{2}$$

$$= \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{ij} - \eta_{im} \\ \eta_{ik} - \eta_{im} \\ \eta_{il} - \eta_{im} \end{pmatrix} = \mathbf{A}^*\,\boldsymbol{\eta}_i^*,$$

where $\mathbf{A}^*$ is the reduced design matrix of $\mathbf{A}$ with the last column of $\mathbf{A}$ removed for identification purpose. The choice between options j and k is determined by the sign of y_{ijk} such that

$$w_{ijk} = \begin{cases} 1, & y_{ijk} \geq 0, \\ 0, & y_{ijk} < 0. \end{cases} \tag{3}$$

Thus, $w_{ijk} = 1$ indicates that option j is preferred to option k for judge i. Accordingly, we have for judge i's complete ranking of r items a corresponding binary response vector $\mathbf{w}_i = (w_{i12}, w_{i13}, \ldots, w_{i(r-1)r})$ for all pairs in the choice set. It is important to note that when we observe complete rankings from n

judges, we only have the data matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$, not the underlying latent responses $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$.

2.2 SEM of ranking data and other measures

Maydeu-Olivares and Böckenholt (2005) showed that estimation and testing of Thurstonian ranking models were made straightforward by formulating the model under an SEM framework as a confirmatory factor model with binary indicators. Consequently, these preference models could be easily estimated using standard SEM softwares and allow for broader array of model possibilities.

For modeling the linear relations of the $\mathbf{y}$ vector of k variables on the $\mathbf{f}$ vector of p latent factors, the following measurement model with mean structure (Jöreskog & Sörbom, 1993) is commonly used such that

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}_y \mathbf{f}_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{\nu}$ is the k-dimensional vector of intercepts, $\boldsymbol{\Lambda}_y$ is the $k \times p$ matrix of factor loadings, and $\boldsymbol{\epsilon}$ is the measurement error associated with $\mathbf{y}$. The covariance matrix of $\boldsymbol{\epsilon}$ is commonly denoted as $\boldsymbol{\Theta}_\epsilon$. For Thurstonian ranking model we have $\boldsymbol{\nu} = \mathbf{0}$, $\boldsymbol{\Lambda}_y = \mathbf{A}$ the $r!/[2!(r-2)!] \times r$ complete comparison design matrix, $\mathbf{f} = \boldsymbol{\eta}$, and $\boldsymbol{\epsilon} = \mathbf{0}$. However, for models with unrestricted covariance structure $\boldsymbol{\Psi}$ of $\boldsymbol{\eta}$, it is preferable to formulate the models using the reduced design matrix $\mathbf{A}^*$ such that $\boldsymbol{\Lambda}_y = \mathbf{A}^*$, the $r!/[2!(r-2)!] \times (r-1)$ matrix for identification purpose. Accordingly, the measurement model is formulated for the latent vector $\boldsymbol{\eta}^*$ instead of $\boldsymbol{\eta}$ and resultantly the model becomes $\mathbf{y}_i = \mathbf{A}^* \boldsymbol{\eta}_i^*$.

The structural part of SEM models is commonly formulated as

$$\mathbf{f}_i = \boldsymbol{\kappa} + \mathbf{B}(\mathbf{f}_i - \boldsymbol{\kappa}) + \boldsymbol{\zeta}_i,$$

where $\boldsymbol{\kappa}$ is the mean vector of $\mathbf{f}$, $\mathbf{B}$ is the matrix containing diagonal elements of zeros and slopes for regressions of latent variables on other latent variables, and $\boldsymbol{\zeta}$ is the vector of residuals and its covariance matrix is usually denoted as $\boldsymbol{\Psi}$. The SEM models for rankings considered in Maydeu-Olivares and Böckenholt (2005) were with no exogenous variables. In other words, the individual differences are simply characterized as deviations from their mean such as $\boldsymbol{\eta}_i = \boldsymbol{\kappa} + \boldsymbol{\zeta}_i$, where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \cdots, \kappa_r)'$ consists of the overall mean utilities and $\mathbf{B} = \mathbf{0}$. The component ζ_{ij} captures the degree to which judge i's latent judgment deviates from κ_j and can be conveniently specified to be normally distributed with $\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$.

As mentioned above, we choose to model $\boldsymbol{\eta}^*$ instead of $\boldsymbol{\eta}$ for identification purpose. Thus, the structural model of the latent variables becomes $\boldsymbol{\eta}^* = \boldsymbol{\kappa}^* + \boldsymbol{\zeta}^*$. The covariance matrix of $\boldsymbol{\zeta}^*$ are $\boldsymbol{\Psi}^* = \mathbf{C}\boldsymbol{\Psi}\mathbf{C}'$ where $\mathbf{C}$ is an $(r-1) \times r$ contrast matrix specified as $\mathbf{C} = [\ \mathbf{I}_{r-1} \quad -\mathbf{1}\ ]$ with the identity matrix of rank $r-1$, $\mathbf{I}_{r-1}$,

and the vector of all ones, $\mathbf{1}$. As a result, the mean and covariance matrices for Thurstonian ranking models are respectively $\boldsymbol{\mu}_y = \mathbf{A}^*\boldsymbol{\kappa}^*$ and $\boldsymbol{\Sigma}_y = \mathbf{A}^*\boldsymbol{\Psi}^*\mathbf{A}^{*\prime}$. Note that we do not directly observe the variables y_{jk}'s. Instead, in our data we have the dichotomization of y_{jk}, i.e., w_{jk}, which is determined by the sign of y_{jk} and indicates which item in the (j, k) pair is preferred within the ranking pattern.

When judge-specific covariates, such as gender or age of the judges, are available, it is useful to account for individual difference at the between-judge level with

$$\boldsymbol{\eta}_i^* = \boldsymbol{\kappa}^* + \boldsymbol{\Gamma}_z\mathbf{z}_i + \boldsymbol{\zeta}_i^*,$$

where the matrix $\boldsymbol{\Gamma}_z$ contains the regression coefficients and $\mathbf{z}_i$ is the judge-specific vector of covariates. The resulting model is also an MIMIC (multiple indicators and multiple causes) model because the latent variables $\boldsymbol{\eta}$ have both effect and cause indicators. Because of the comparative nature of the data, only the main effects of covariates on the differences between item utilities rather than on individual item utilities can be identified. Instead of fitting the model to the covariance or correlation matrix of $(\mathbf{y}, \mathbf{z})$, it is a common practice to consider the conditional distribution of $\mathbf{y}$ given the covariates $\mathbf{z}$.

When data other than rankings are available from the same judges, it might be of interest to see how their utilities of the choice items are related to or influenced by their personality traits measured by the additional data. For example, one might be interested in how judges' quality of life relates to their ranking of a number of pension plans, or of a group of presidential candidates. In such cases, additional factors which presumably correlate with the item utilities are introduced into the measurement model such that

$$\mathbf{x}_i = \boldsymbol{\Lambda}_x\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \tag{4}$$

where $\boldsymbol{\xi}$ is the vector of factors of interest and $\boldsymbol{\Lambda}_x$ represents the matrix of factor loadings of the observed response vector $\mathbf{x}$ on the vector of factors $\boldsymbol{\xi}$. With the inclusion of the additional latent vector $\boldsymbol{\xi}$ which is assumed to have a direct effect on the option utilities, the structural part of the model becomes

$$\begin{pmatrix} \boldsymbol{\eta}^* \\ \boldsymbol{\xi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\kappa}^* \\ \boldsymbol{\kappa}_\xi \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \boldsymbol{\Gamma}_\xi \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}^* - \boldsymbol{\kappa}^* \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_{\eta^*} \\ \boldsymbol{\zeta}_\xi \end{pmatrix}, \tag{5}$$

where the mean and covariance matrices of $\boldsymbol{\xi}$ are $\boldsymbol{\kappa}_\xi$ and $\boldsymbol{\Phi}$. However, for simplicity, here we are not particularly interested in the mean structure of $\boldsymbol{\xi}$ and therefore we set $\boldsymbol{\kappa}_\xi = \mathbf{0}$. In addition, we set $\boldsymbol{\Phi}$ to be the identity matrix $\mathbf{I}$ for identification purpose. When all the response variables in $\mathbf{x}$ are categorial, threshold parameters used for discretization also need to be estimated. The direct effect of $\boldsymbol{\xi}$ on item utility vector $\boldsymbol{\eta}^*$ will help understanding the plausible causes of individual differences in judges' utilities.

By combining the above approaches, i.e., using both observed covariates and latent factor, to account for individual differences, we arrive at the resultant structural part of the model that

$$
\begin{pmatrix} \boldsymbol{\eta}^* \\ \boldsymbol{\xi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\kappa}^* \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \boldsymbol{\Gamma}_\xi \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}^* - \boldsymbol{\kappa}^* \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_z \\ \mathbf{0} \end{pmatrix} \mathbf{z}_i + \begin{pmatrix} \boldsymbol{\zeta}_{\eta^*} \\ \boldsymbol{\zeta}_\xi \end{pmatrix}, \quad (6)
$$

where the covariance matrices for $\boldsymbol{\zeta}_{\eta^*}$ and $\boldsymbol{\zeta}_\xi$ are respectively $\boldsymbol{\Psi}^*$ and $\boldsymbol{\Phi} = \mathbf{I}$. Under this model, the conditional expectation and variance of $(\mathbf{y}, \mathbf{x})$ given $\mathbf{z}$ are

$$
\mathrm{E}(\mathbf{y}, \mathbf{x}|\mathbf{z}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1} \begin{pmatrix} \boldsymbol{\kappa}^* + \boldsymbol{\Gamma}_z \mathbf{z}_i \\ \mathbf{0} \end{pmatrix}, \quad (7)
$$

$$
\mathrm{Var}(\mathbf{y}, \mathbf{x}|\mathbf{z}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \quad (8)
$$

where $\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_x \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \boldsymbol{\Gamma}_\xi \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, $\boldsymbol{\Psi} = \mathrm{Cov} \begin{pmatrix} \boldsymbol{\zeta}_{\eta^*} \\ \boldsymbol{\zeta}_\xi \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Psi}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$, and $\boldsymbol{\Theta} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_\delta \end{pmatrix}$.

In summary, let $\boldsymbol{\theta}$ denote the set of all the parameters in the above-mentioned SEM model, we have $\boldsymbol{\theta} = \{\boldsymbol{\kappa}^*, \boldsymbol{\Gamma}_z, \boldsymbol{\Lambda}_x, \boldsymbol{\Gamma}_\xi, \boldsymbol{\Psi}^*, \boldsymbol{\Theta}_\delta\}$. By obtaining and interpreting the estimates $\hat{\boldsymbol{\theta}}$ for the parameters in the SEM model, we would be able to gain some understanding of individual differences in the utilities the judges assigned to the options in a ranking task and how these utilities are related to either the covariates or some personality traits of the judges.

3. Data Analysis

3.1 The data: detroit area study

The data were taken from the inter-university consortium for political and social research (ICPSR) data archives and collected by Alwin (1997) investigating the social change in religion and child rearing in the Detroit Area Study. For the survey, respondents from three counties in the Detroit area were queried about their work, health, marriage and family, finances, political views, religion, and child rearing. For current purposes, we specifically choose their responses to a preference ranking task and a number of questions which we hypothesized of forming a single construct, in addition to some background variables of the respondents. The chosen ranking question and its choice set in the original study are:

Question: "While we're talking about children, please tell me which of the following things you would pick as more important for children to learn to prepare

them for life. If you had to choose, is it more important for children to learn"
(Alwin, 1997).

1. To obey

2. To think for themselves

3. To work hard

4. To help others when they need help

Note that in the original study, there was another choice item "To be well-liked
and popular". However, this item caused estimation difficulty for making some
of the two-way table with zero frequencies and therefore was dropped from the
subsequent analysis. In other word, we extracted the respondents' preference
ranking of the above four items from their responses to the five items. Moreover,
the items for the single factor which presumably influences the utility evaluations
are,

x_1 : Is gambling wrong?

x_2 : Is divorce wrong?

x_3 : Is homosexuality wrong?

x_4 : Is legal abortion allowed for married couples who do not want any more
children?

x_5 : Is assisted suicide allowed?

For the first three items, their response categories were ordinal (1=always; 2=usu-
ally; 3=sometimes; 4=never) whereas the last two contained binary outcomes
(1=yes, allowed; 2=no, not allowed). In addition, gender and marital status were
treated as known sources of individual differences and were incorporated as co-
variates in the model. More specifically, three groups were defined for marital
status as *married*, *once married* (either widowed, divorced, or separated), and
never married. The proportion of females in the total sample was .63 and the
proportions of *married*, *once married*, and *never married* were respectively .57,
.25, and .18.

   Tables 1 and 2 showed that gender and marital status were indeed inducing
heterogeneity in the sample. The rank ordering implied by the mean ranks were
different for males (help others $\succ$ obey $\succ$ think-for-self $\succ$ work-hard) and for
females (help others $\succ$ work-hard $\succ$ think-for-self $\succ$ obey). For marital status, it
can be seen that people who are married or once married are different from those

who are never married only in their preference between "to obey" and "to think for themselves". More specifically, the former seemingly preferred kids to obey than to think for themselves whereas the latter on average showed the opposite.

Table 1: Summary of ranks by gender

Gender	obey	think-for-self	work-hard	help-others
Female				
M(SD)	2.95(1.10)	2.82(.99)	2.52(1.04)	1.70(.89)
Male				
M(SD)	2.61(1.10)	2.79(1.06)	2.84(1.04)	1.77(.94)

M=Mean of ranks; SD=Standard Deviation of ranks

Table 2: Summary of ranks by marital status

Marital Status	obey	think-for-self	work-hard	help-others
Married				
M(SD)	2.78(1.10)	2.85(1.01)	2.69(1.03)	1.68(.91)
Once-Married				
M(SD)	2.76(1.16)	2.80(1.01)	2.58(1.11)	1.86(.94)
Never-Married				
M(SD)	3.06(1.08)	2.71(1.05)	2.56(1.05)	1.69(.83)

M=Mean of ranks; SD=Standard Deviation of ranks

3.2 Estimation and testing

The parameter estimates of the proposed choice models can be obtained by limited-information (LI) maximum likelihood method. Unlike the full-information maximum likelihood methods which require multidimensional integrals to be evaluated numerically and consequently the computation could be cumbersome when r is large. LI algorithms are computationally less demanding and considerably faster and therefore are better suited to effectively estimate the model parameters. The suitability of LI method for the analysis of ranking data was first recognized by Maydeu-Olivares (1999, 2001).

LI estimation for categorical data is usually carried out in two stages. Here, the thresholds and the tetrachoric correlations between the binary responses in $\mathbf{W}$ are estimated first. At the second stage, model parameters are estimated by minimizing the weighted or unweighted least squares of the difference between the estimates obtained from the sample at the first stage and their model-based specification. Let $\boldsymbol{\rho}$ and $\hat{\boldsymbol{\rho}}$ denote the vector of all thresholds and tetrachoric

correlations between the binary responses in $\mathbf{W}$ and its corresponding parameter estimates at the first stage, respectively. At the second stage, the estimates $\hat{\boldsymbol{\theta}}$ of the parameter set $\boldsymbol{\theta}$ consisting of all parameters in the SEM model are obtained to satisfy restrictions imposed on $\boldsymbol{\rho}$ by equation (6) so that the following least squares function with the weight matrix $\mathbf{V}_\rho$ is minimized:

$$F = \left[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\hat{\boldsymbol{\theta}})\right]' \mathbf{V}_\rho \left[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\hat{\boldsymbol{\theta}})\right].$$

Two choices of the weight matrix $\mathbf{V}_\rho$ based on functions of the estimated asymptotic covariance matrix $\hat{\boldsymbol{\Xi}}$ of $\hat{\boldsymbol{\rho}}$ have been proposed which can result in consistency and asymptotical normality for the parameter estimators (Muthén, 1978, 1984; Muthén, du Toit & Spisic, 1997). They are the weighted least squares (WLS) and the diagonally weighted least squares (DWLS) approaches using respectively $\mathbf{V}_\rho = \hat{\boldsymbol{\Xi}}^{-1}$ and $\mathbf{V}_\rho = \left[\mathrm{Diag}(\hat{\boldsymbol{\Xi}})\right]^{-1}$. For our analysis, the DWLS approach is used because it has been shown that larger samples are needed for WLS than DWLS to obtain reasonable parameter estimates and standard errors (Muthén, 1993).

Goodness-of-fit tests based on the statistic $T = n\hat{F}$ are available for both WLS and DWLS, where n is the sample size. Satorra and Bentler (1994) and Muthén (1993) showed that T, though not asymptotically chi-square distributed itself under DWLS, can be used for testing after scaling T by its asymptotic mean or by its asymptotic mean and variance. The Satorra-Bentler (SB) mean adjusted statistic, $\bar{T} = T/\hat{c}$ where $\hat{c}$ is the scaling factor, can be used for both the assessment of the overall fit of a model and fit comparisons of nested models under DWLS (Satorra and Bentler, 1994). We note that in most cases it would be incorrect to compute the difference between the SB-scaled statistics of two nested models and to test the obtained value against a χ^2- variate with the degrees of freedom equal to the difference between the number of independent parameters estimated under the two nested models (Satorra & Bentler, 2001). Instead, the scaled χ^2-difference test statistic $\bar{T}_d$ should be employed for comparing two nested models (Satorra & Bentler, 2001). Let M_0 and M_1 denote the two nested model where M_1 is less restricted and r_0 and r_1 be their associated degrees of freedom of the goodness-of-fit statistics T_0 and T_1, respectively. The scaled χ^2-difference test statistic, $\bar{T}_d$, can then be obtained as

$$\bar{T}_d := T_d/\bar{c}_d, \quad \bar{c}_d := (r_0\hat{c}_0 - r_1\hat{c}_1)/(r_0 - r_1),$$

where $\hat{c}_0$ and $\hat{c}_1$ are the scaling factors of T_0 and T_1 respectively, and $T_d = T_0 - T_1$ with $T_0 = \hat{c}_0\bar{T}_0$ and $T_1 = \hat{c}_1\bar{T}_1$. Under the null hypothesis, $\bar{T}_d$ is asymptotically distributed as a χ^2-variate with degrees of freedom $r_0 - r_1$.

The total number of degrees of freedom for the chi-squares statistics reported in the SEM software Mplus is computed as the sum of the number of thresholds

(for both binary or paired comparisons and additional categorical factor indicators), the product of the number of independent columns of covariates and the number of categorical responses, and all the polychoric (tetrachoric) correlations. For example, consider the model of paired comparisons of four items with two covariates, gender and age, and an additional factor which presumably associates with the item utilities and contains five binary indicators, then its total degrees of freedom is

$$\mathrm{df}_T = [4!/(2!2!)+5]+2\times[4!/(2!2!)+5]+\{[4!/(2!2!)+5][4!/(2!2!)+5-1]\}/2 = 88.$$

Maydeu-Olivares and Böckenholt (2005) pointed out that because rankings give rise to only a subset of all possible paired comparison binary outcomes and an adjustment on the number of degrees of freedom is therefore required when modeling ranking data are analyzed using SEM models for dichotomous items. More specifically, when ranking data are transformed into binary data and analyzed alone, there will be $d = [r(r-1)(r-2)]/6$ redundancies among the thresholds and tetrachoric correlations estimated from the binary variables (Maydeu-Olivares, 1999). Therefore, we propose subtracting the constant d from the degrees of freedom reported for the extended SEM model for paired comparisons or binary response. However, the goodness-of-fit of the specified model can also be assessed using other fit indices such as the Comparative-Fit-Indices (CFI), Tucker-Lewis-Indices (TLI), Root Mean Square Error Of Approximation (RMSEA) and Weighted Root Mean Square Residual (WRMR). Although the calculation of the above fit indices also require an adjustment on the degrees of freedom for the model, the effect of a small difference in the degrees of freedom should have less impact on those fit statistics than on the chi-squared statistics. Yu (2002) suggested, based on a series of simulation studies on CFA and MIMIC models, that to find suitable cutoff criteria and to understand how fit indices perform under a certain type of models with a certain sample size, substantive researchers might want to conduct their own simulation studies and summarized that CFI $\geq$ .96, RMSEA ≤ 0.05, and WRMR ≤ 1.0 can be indications of good models with binary outcomes at sample sizes larger than 250.

3.3 Identification constraints

Because of the discrete nature of ranking data, some identification constraints are needed for the estimation of their mean and covariance parameters, $\boldsymbol{\kappa}$ and $\boldsymbol{\Psi}$, even in the simple case with only the comparative judgment data. Here we simply use the utility difference where the last item is chosen as the baseline item and the parameters of interest are $\mathbf{C}\boldsymbol{\kappa}$ and $\mathbf{C}\boldsymbol{\Psi}\mathbf{C}'$, where $\mathbf{C}$ is the contrast matrix of which each column indicating the comparison between an item to the

last. That is, as noted before, the parameters to be estimated are

$$\mathbf{C}\boldsymbol{\kappa} = \boldsymbol{\kappa}^* = \left(\begin{array}{c} \kappa_1^* \\ \kappa_2^* \\ \kappa_3^* \end{array} \right), \text{ and } \mathbf{C}\boldsymbol{\Psi}\mathbf{C}' = \boldsymbol{\Psi}^* = \left(\begin{array}{ccc} \psi_{11}^* & \psi_{21}^* & \psi_{31}^* \\ \psi_{21}^* & \psi_{22}^* & \psi_{32}^* \\ \psi_{31}^* & \psi_{32}^* & \psi_{33}^* \end{array} \right),$$

where ψ_{11}^* is further fixed to 1 for identification purpose.

When other measures in addition to rankings are included in the analysis, constraints necessary for identification of the whole model often need some investigation. For the case that not all additional variables are continuous, the residual variances for continuous latent response variables of observed categorical dependent variables are often treated as parameters to be estimated. However, the number of parameters in this case will result in under-identification because we are fitting covariance structure to those categorical dependent variables while only correlations between them are identified. One possible solution is to impose sufficient restrictions on the residual variances to ensure identification of the model parameters, without putting additional constraints on them. Cudeck (1989) discussed the interrelated problems associated with those models where covariance structure are fitted to sample correlations and concluded that, for the class of scale-invariant models (Browne, 1982), applying a covariance structure to a matrix of correlations will not modify the models under consideration. Consequently, because our intended single factor model for the five indicators of personality trait in addition to rankings is scale-invariant, the further constraint of $\boldsymbol{\Theta}_\delta = \mathbf{I} - \text{diag}(\boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x')$ could be used for identification of the whole model. However, when the intended model is not scale-invariant, the constraints necessary to ensure identification without imposing any additional restrictions on the model become less straightforward.

3.4 Results

The estimation was carried out with the SEM software Mplus. We first examined the total number of degrees of freedom for the model under consideration. Because the data contain ranking of four items, two binary indicators and three four-category indicators and two covariates (one is binary and the other with three levels), the total number of degrees of freedom reported in Mplus would be 105 and consequently the corrected degrees of freedom for our model for ranking will be $\text{df}_T = 105 - d = 105 - [(4)(3)(2)]/6 = 101$.

We first considered a model with simply the ranking data (M_0 in Table 3). With the corrected degrees of freedom of 9 in this case, the values of the selected fit indexes $\chi^2(9) = 33.784$, $p = .0001$, CFI and TLI were respectively .977 and .974, RMSEA=.078 and WRMR=.731. Three indices, CFI, TLI, and WRMR indicated that M_0 yielded reasonable fit to the ranking data whereas χ^2 statistic

and RMSEA were not favorable. That is, a close fit or approximation of the
model to the ranking data might be slightly doubtful. For M_0, it was assumed
that the data were from a homogeneous group of judges. To further incorporate
the possible sources of individual difference to account for heterogeneity, we tried
next to include both covariates, gender and marital status, and an additional
factor which presumably influenced judges' rankings into modeling the choice
behavior of the judges.

Table 3: Goodness-of-fit statistics for icpsr ranking with additional factor and
covariates

		Goodness-of-fit				
Models	df	scaled χ^2 ($\bar{T}$)	p	CFI/TLI	RMSEA	covariates
M_0	9	33.784	.0001	.977/.974	.078	None
M_1	65	88.717	.0270	.989/.985	.033	η_1^* (G, M), η_2^* (G, M), η_3^* (G,M)
M_2	67	86.541	.0544	.991/.989	.029	η_1^* (G, M), η_2^* (G,M), η_3^* (G)
M_3	68	91.039	.0326	.989/.986	.032	η_1^* (G, M), η_2^* (M), η_3^* (G)
M_4	69	88.134	.0601	.991/.989	.028	η_1^* (G, M), η_2^* (G), η_3^* (G)
M_5	70	92.525	.0371	.989/.987	.031	η_1^* (G, M), η_3^* (G)
M_6	49	73.140	.0143	.988/.985	.038	η_1^* (G), η_2^* (G) η_3^* (G)
M_7	50	76.769	.0088	.986/.983	.040	η_1^* (G), η_3^* (G)

G: Gender; M: marital status

A total of another seven different models were fit to the ranking data which
differed in their choice of nonzero covariate effects, with the inclusion of both
an additional factor and the covariates. Table 3 provided a summary of the
goodness-of-fit statistics obtained from these analyses. Model specifications were
given in the last column of Table 3 labeled 'covariates'.

M_1 and M_6 served as benchmark models for cases where respectively both
covariates (gender and marital status) or only gender were included in the models.
For models with the same number of covariates, such as M_1 to M_5, they were
considered as nested model and therefore could be tested using the scaled SB χ^2
difference statistic $\bar{T}_d$. As a result, we concluded that M_4, among the models with
both covariates, provided the most parsimonious representation of the data with
$\bar{T} = 88.134$ and $p = .0601$. Moreover, M_6 and M_7 were models with only gender
as the sole covariate and might not be considered as nested model with M_4 and
therefore the scaled SB χ^2 difference test might not be directly applicable in this
case. However, based on the p-value associated with either model we found that
M_4 would be considered preferable to both M_6 and M_7.

Model 4 (M_4) was depicted graphically in Figure 1. The parameter estimates
for M_4 were reported in Table 4. The column labeled "Standardized" contained

the parameter estimates standardized with respect to the variances of $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{z}$. The results in the top part of the table concerned the mean structure and the effect of gender and marital status on the perceived utilities associated with the items. For example, $\hat{\mathbf{\Gamma}}_G = (\ .218\quad .106\quad .124\ )'$ indicated that males (gender=1) had a mean utility difference between "to obey" and "to help others" which was $(2)(.218)=.436$ higher than females (gender=-1). The difference between males and females on their mean utilities were different for different choice items (.212 for η_2^* and .248 for η_3^*). However, for martial status, it only made a difference for judges' associated utilities on η_1^*, not on the other two. That is, the effect of being *married, once married,* and *never married* on the mean of η_1^* were respectively -.102, .167 and -.065. In other words, although on average "to help others" was ranked more important than "to obey" by the judges, married people felt more so than those who were never married. Those who were once married, though on average still considered "to help others" to be more important than "to obey", the difference in their degrees of importance was much smaller while compared to the other two groups. Moreover, using the unstandardized estimates of the intercept and the effects of gender and martial status, the mean differences of the latent utilities η_1^*, η_2^*, and η_3^* for married males were estimated as follows:

$$
\hat{\boldsymbol{\kappa}}^* = \begin{pmatrix} \hat{\kappa}_1^* \\ \hat{\kappa}_2^* \\ \hat{\kappa}_3^* \end{pmatrix} = \begin{pmatrix} -.734 \\ -.272 \\ -.287 \end{pmatrix} + (1)\begin{pmatrix} .218 \\ .106 \\ .124 \end{pmatrix} + (1)\begin{pmatrix} -.102 \\ 0 \\ 0 \end{pmatrix} + (0)\begin{pmatrix} .167 \\ 0 \\ 0 \end{pmatrix}
$$
$$
= \begin{pmatrix} -.618 \\ -.166 \\ -.163 \end{pmatrix}.
$$

The mean differences for other groups can be similarly estimated. Parameter estimates reported in the lower part of Table 4 concerned the covariance structure of M_4. The largest unstandardized (standardized) coefficient for covariances between the single factor and the utilities showed that the factor was negatively correlated with η_1^*, their judgment of "to obey" over "to help others" with a standardized $\hat{\gamma}_1 = -.480$. That is, the lower the judge scored on the factor, the more he or she preferred "to obey" over "to help others". Based on the five indicators and their loadings on the single factor, $\hat{\mathbf{\Lambda}}_x = (\ .641\quad .550\quad .692\quad -.829\quad -.690\ )'$, we conveniently defined the factor as the level of conservativeness and the lower the factor score, the higher level of conservativeness. In other words, it implied that more conservative a judge was, he or she would tend to choose "to obey" more in comparison to "to help others". However, we noted that based on the estimated values of $\boldsymbol{\kappa}$, on average, "to help others" was the most preferred choice.

The correlation between η_1^* and η_2^*, after accounting for the covariate effect and the variability attributed from the factor of conservativeness, appeared to be
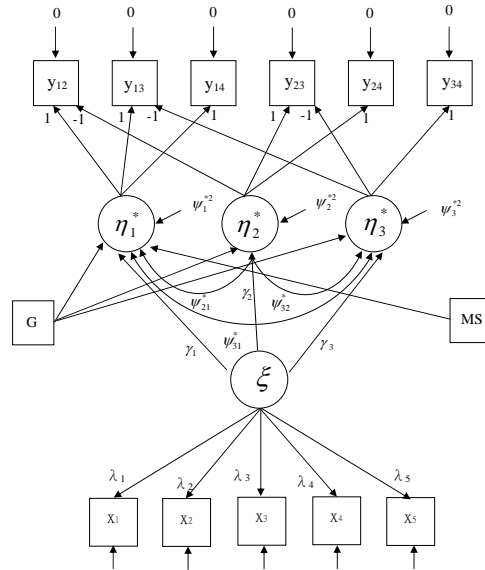
Figure 1: Covariance structure of model M_4 for Detroit Area Study with rankings (y_{12} to y_{34}), responses to five attitude questions (x_1 to x_5), and two covariates of the respondents (G: gender; MS: martial status).

very small with an estimated value of -.061 whereas those between η_3^* and either η_1^* and η_2^* were considerably larger with estimated values of .310 and .422 respectively. That is, the assignment of utilities to option "to work hard" were positively correlated with those to either "to obey" or "to think of themselves", in comparison to "to help others".

4. Conclusion

In this paper, we extend the SEM approach by Maydeu-Olivares and Böckenholt (2005) to model simultaneously ranking data and other types of categorical data, in addition to the inclusion of covariates. An detailed analysis of Detroit Area Study has shown the usefulness of the proposed approach in providing a better understanding of individual difference in their choice behavior. The advantage of formulating the Thurstonian choice models as structural equation models indeed facilitate possible extensions beyond merely option scaling. More studies are yet necessary to evaluate the validity and performance of S-B statistics in comparing nested or non-nested models under different setting of sample size, model complexity, and especially for models with different numbers of covariates. However, this outlook for future work should not distract from the fact that, taking advantage of the advances in SEM modeling literature, a wider range of applications would now become more accessible for analyzing ranking data.

Table 4: Parameter Estimates for model M_4

Parameter		Unstandardized	SE	Standardized
intercept ($\boldsymbol{\kappa^*}$)	κ_1^*	-.734	.093	-.717
	κ_2^*	-.272	.078	-.317
	κ_3^*	-.287	.067	-.386
effect of gender				
$\boldsymbol{\Gamma}_G$	γ_{G,η_1^*}	.218	.081	.206
	γ_{G,η_2^*}	.106	.067	.119
	γ_{G,η_3^*}	.124	.057	.161
effect of marital status				
$\boldsymbol{\Gamma}_{MS}$	γ_{MS_1,η_1^*}	-.102	.101	-.077
	γ_{MS_2,η_1^*}	.167	.119	.106
$\boldsymbol{\Psi}^*$	ψ_1^{*2}	1.00*		.954
	ψ_2^{*2}	.709	.166	.986
	ψ_3^{*2}	.535	.122	.974
	ψ_{12}^*	-.054	.079	-.061
	ψ_{13}^*	.236	.079	.310
	ψ_{23}^*	.269	.077	.422
$\boldsymbol{\Gamma}_\xi$	γ_1	-.491	.073	-.480
	γ_2	.169	.080	.197
	γ_3	-.084	.059	-.113
factor loadings				
$\boldsymbol{\Lambda}_x$	λ_1	.641	.053	.641
	λ_2	.550	.056	.550
	λ_3	.692	.060	.692
	λ_4	-.829	.052	-.829
	λ_5	-.690	.064	-.690
measurement errors				
$\boldsymbol{\Theta}_\delta$	θ_{δ_1}	.589	.068	.589
	θ_{δ_2}	.698	.062	.698
	θ_{δ_3}	.522	.084	.522
	θ_{δ_4}	.312	.087	.312
	θ_{δ_5}	.525	.088	.525

References

Alwin, D. (1997). Detroit Area Study: Social change in religion and child rearing. Inter-university Consortium for Political and Social Research (ICPSR0-4120). Ann Arbor, MI.

Bock, R. D. (1958). Remarks on the test of significance for the method of paired comparisons. *Psychometrika* **23**, 323-334.

Böckenholt, U. and Tsai, R. (2001). Individual differences in paired comparison data. *British Journal of Mathematical and Statistical Psychology* **54**, 265-277.

Browne, M. W. (1982). Covariance structures. In D.M.Hawkins (Ed.), *Topics in Applied Multivariate Analysis* (pp. 72-141). Cambridge University Press, Cambridge.

Cudeck, R. (1989). Analysis of correlation matrices using covariancec structure models. *Psychological Bulletin* **105**, 317-327.

Furuya, H., Fujimoto, K., Wakuda, A. and Nakano, Y. (2005). The influence of total and directional energy of late sound on listener envelopment. *Acoustical Science and Technologym* **26**, 208-211.

Jöreskog, K. G. and Sörbom, D. (1993). *LISREL* 8: *User's reference guide.* *Scientific Software*, Chicago.

Kojima, T. and Kusumi, T. (2006). The effect of an extra object on the linguistic apprehension of the spatial relationship between two objects. *Spatial Cognition and Computation* **6**, 145-160.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika* **64**, 325-340.

Maydeu-Olivares, A. (2002). Limited information estimation and testing of Thurstonian models for preference data. *Mathematical Scoial Sciences* **43**, 467-483.

Maydeu-Olivares, A. and Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods* **10**, 285-304.

McFadden, D. (2001). Economic Choices. *American Economic Review* **91**, 351-378.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika* **43**, 551-560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49**, 115-132.

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. In: Bollen, K.A. and Long, J.S. (Eds.), *Testing Structural Equation Models* (pp. 205-234). Sage, Newbury Park.

Muthén, B., du Toit, S.H.C. and Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript, College of Education, University of California, Los Angeles.

Prietoa, L. B. and Alonsoa, J. J. (2000). Exploring health preferences in sociodemographic and health related groups through the paired comparison of the items of the Nottingham Health Profile. *Journal of Epidemiology and Community Health* **54**, 537-543.

Satorra, A. and Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In: A. von Eye and C.C. Clogg (Eds.), *Latent variable analysis*: *Applications to developmental research* (pp. 399-419). Sage, Thousand Oaks

Satorra, A. and Bentler, P. M. (2001). A scaled difference chisquare test statistic for moment structure analysis. *Psychometrika* **66**, 507-514.

Takane, Y. (1987). Analysis of covariance structures and probabilistic binary choice data. *Cognition and Communication* **20**, 45-62.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review* **34**, 273-286.

Train, K. (2003). *Discrete Choice Methods with Simulations.* MIT Press, Cambridge.

Yao, G. and Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology* **52**, 79-92.

Yu, P. L. H. (2000). Bayesian analysis of order-statistics models for ranking data. *Psychometrika* **65**, 281-299.

Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes.* Doctoral dissertation, University of California, Los Angeles.

Rung-Ching Tsai
Department of Mathematics
National Taiwan Normal University
No. 88, Sec. 4, Ting-Chou Rd,
Taipei 116, Taiwan
rtsai@math.ntnu.edu.tw

Ulf Böckenholt
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL, USA 60208
u-bockenholt@kellogg.northwestern.edu